Network Reconstruction and Analysis: TP1

Mathieu Pont and Lucas Rodrigues Pereira

December 2019

Abstract

We discuss the correlation and partial correlation of transcription factors (TFs) implicated in the differentiation of primitive streak cells at the time of endothelial and hematopoietic differentiation. We show that, in general, TFs associated with a same differentiation are correlated whereas those associated with different ones are negatively correlated.

1 Introduction

In the context of the Master's Degree on Machine Learning for Data Science, at the University of Paris (Descartes), we have been given the task to discuss correlation between transcription factors, as part of the "Network Reconstruction and Analysis" course.

We used the *hematoData* dataset representing 33 transcription factors involved in the differentiation of 3934 cells. The cells are either differentiated to endothelial or hematopoietic cells.

2 Correlation Network

We used the Pearson correlation to make an adjacency matrix between the 33 transcription factors (TFs). If all the values are kept the graph is complete making it unreadable. Therefore, we kept only the 5 highest positive edges and the 4 negative ones for each node.

The coloration for vertices is the same as [Verny et al., 2017]. It indicates to which group the node belongs. The red color is for the hematopoietic TFs. Endothelial are colored in violet. Those common to both are in blue and finally the unclassified in gray.

Edges are colored according to their values, negative correlation are in red whereas positive ones in blue. The width of each edge is proportional to its absolute value.

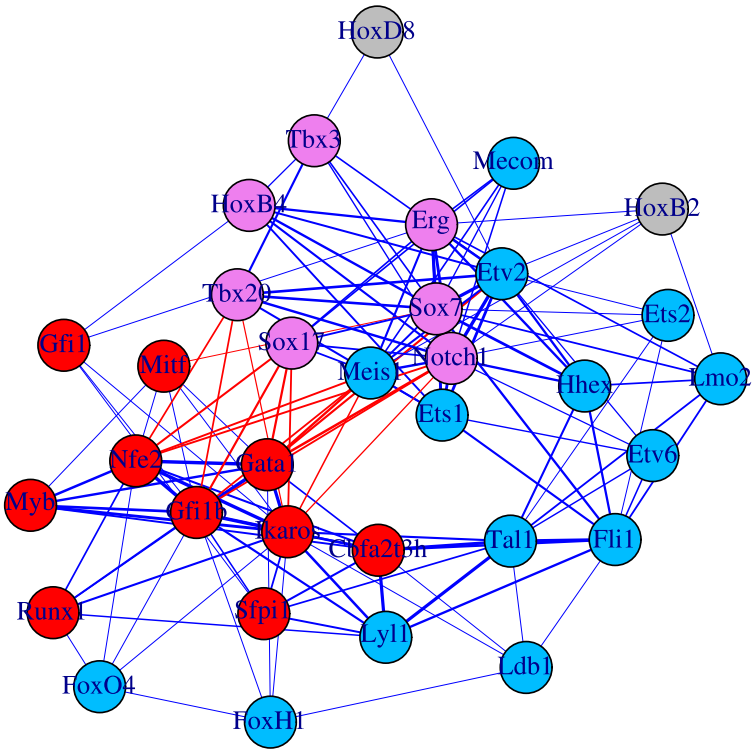


Figure 1: Correlation Network.

It is interesting to see that, only with correlation, each TFs groups are well separated, especially for the hematopoietic and endothelial groups. The one common to the last two are widespread along them, which makes sense since it is related to them.

The negative correlation (red edges, representing repressions) are especially between the hematopoietic group and the endothelial one. This result was expected since they are two different possible differentiation of the cells with different goals, therefore these cells should not have TFs associated to both groups making their correlation negative.

We also found known relations between TFs. For example, like found in [Ferreirós-Vidal et al., 2013] we also found a relation between Ikaros and Gfi1b and one between Ikaros and Lyl1. Like in [Eppig et al., 2014] we found a relation between Tal1 and Fli1 and one between Tal1 and Lmo2. And like in [Moignard et al., 2015] we found a relation between HoxB4 and Erg and one between Sox7 and Erg.

Finally, our network is quite similar to the one in [Verny et al., 2017]. The repressions are more

or less between the same group of nodes. The pairwise relation of each pair of nodes are quite similar too, we can easily found same relations in both network. However, we did not have the direction of edges and some special relations as in the paper.

3 Partial Correlation Network

For this part, we have used the inverse of the covariance matrix (Σ^{-1}) to extract partial pairwise correlations. One peculiarity of this method is that it fixes all other variables. Thus, it tends to reduce partial correlations the greater the number of fixed variables were.

This method has captured similar correlations, as compared to the Pearson correlation. The repressions, as expected, are mainly between Endothelial and Hematopoietic cells. However, some new unexpected negative partial correlations have appeared, such as between Etv2 (Primitive Streak) and Runx1 (Hematopoietic).

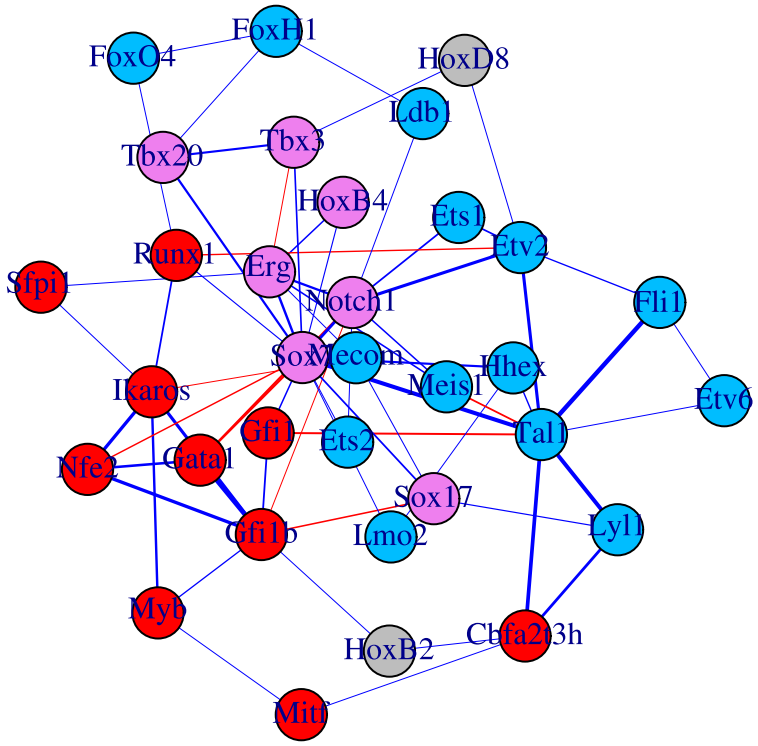
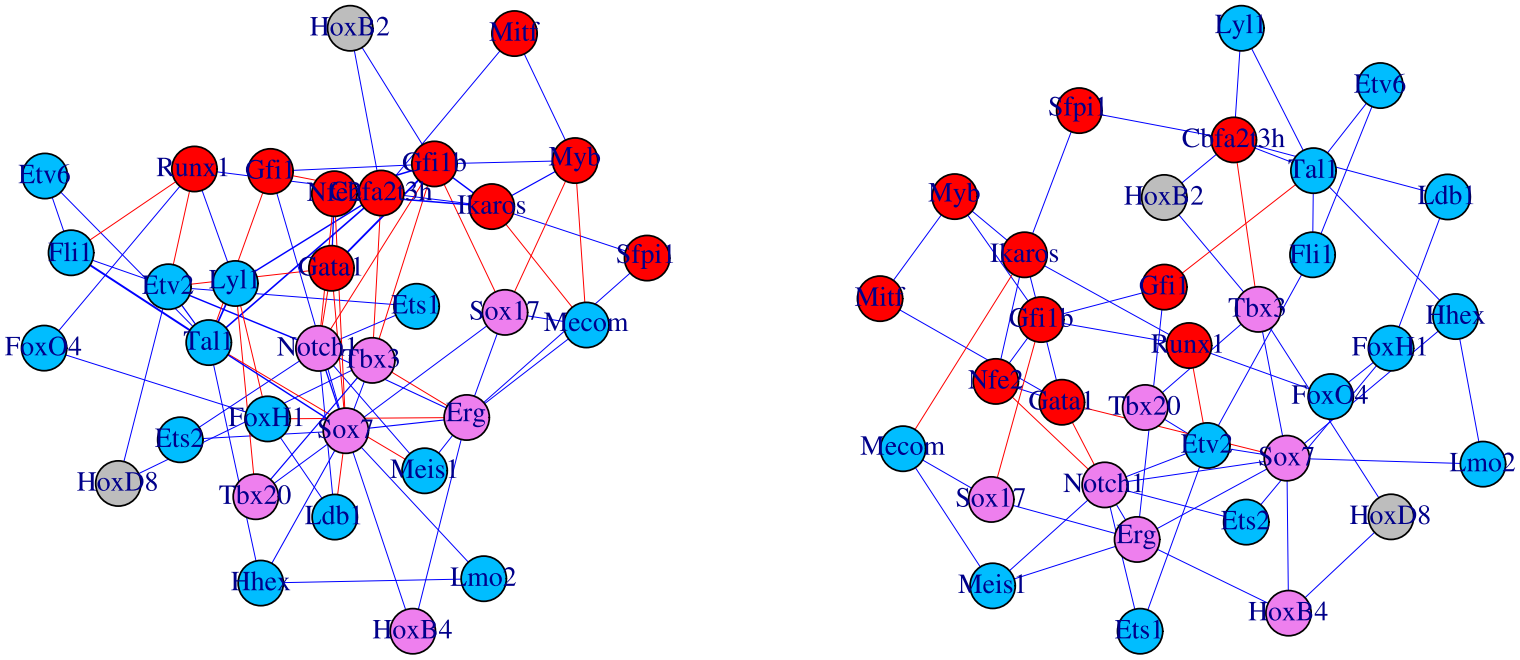


Figure 2: Partial Correlation Network.

In order to invert the covariance matrix, it is required to be invertible, which cannot be done unless the determinant is not null. One usual trick to make it invertible is to perform a regularization by adding a small value λ to the covariance matrix diagonal. However, it seems to weaken the partial correlations. Moreover, a negative correlation between Tal1 and Meis1, both from the Primitive Streak class, has appeared.



(a) Partial Correlation Network with $\lambda = 0.2$.

(b) Partial Correlation Network with $\lambda = 1$.

Figure 3: Partial Correlation Network for different λ.

As seen on Figure 3a some negative correlations appeared compared to the "basic" correlation network. For example between Lyl1 and Tbx20, between Lyl1 and FoxH1 or between Sox7 and Ldb1.

We still have a good network compared to the one in [Verny et al., 2017]. However, we have lost some known relations (for example the one between Tal1 et Lmo2) but we found new ones compared to the "basic" network.

4 Conclusion

We have tested two different approaches to reconstruct the correlation graphs representing the relation among variables, in this case: cell expression data at the time of endothelial and hematopoietic differentiation from the primitive streak cells of the mouse early embryo.

First we have used the Pearson correlation, which have performed well, as compared to the original study by [Verny et al., 2017.]. Then, we have used partial correlations, also pairwise, but now taking into consideration the impact of the other variables. New negative correlations have been exposed, even though the overall correlation between the genes have been kept relatively the same.

Finally, to assess our results, some known relations in the literature was also found in our networks.

References

[Eppig et al., 2014] Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., and Group, T. M. G. D. (2014). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Research*, 43(D1):D726–D736.

[Ferreirós-Vidal et al., 2013] Ferreirós-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B. S., Smale, S. T., Spivakov, M., Srivastava, P., Petretto, E., Fisher, A. G., and Merkenschlager, M. (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B–cell differentiation. *Blood*, 121(10):1769–1782.

[Moignard et al., 2015] Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., Buettner, F., Macaulay, I. C., Jawaid, W., Diamanti, E., Nishikawa, S.-I., Piterman, N., Kouskoff, V., Theis, F. J., Fisher, J., and Göttgens, B. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature Biotechnology*, 33(3):269–276.

[Verny et al., 2017] Verny, L., Sella, N., Affeldt, S., Singh, P., and Isambert, H. (2017). Learning causal networks with latent variables from multivariate information in genomic data. *PLOS Computational Biology*, 13:e1005662.